E-FILED on 8/31/2011

IN THE UNITED STATES DISTRICT COURT

FOR THE NORTHERN DISTRICT OF CALIFORNIA

SAN JOSE DIVISION

IN RE GOOGLE LITIGATION	No. C-08-03172 RMW CLAIM CONSTRUCTION ORDER **[Re Docket No. 246]**

On November 21, 2007, Software Rights Archive, LLC ("SRA") sued defendants Google Inc., Yahoo! Inc., IAC Search & Media Inc., Lycos, Inc. and AOL, LLC (collectively "defendants") in the Eastern District of Texas. SRA asserts that defendants infringe United States Patent No. 5,544,352 ("'352 Patent"), United States Patent No. 5,832,494 ("'494 Patent"), and United States Patent No. 6,233,571 ("'571 Patent"). On July 21, 2008, defendants filed a declaratory judgment action in this court, alleging non-infringement and invalidity of the same three patents. The Texas action was transferred to this court on July 22, 2010 and then consolidated with the declaratory judgment action. The parties seek construction of claim terms in the asserted patents. The court held a claim construction hearing on April 27, 2011. Per the court's request, the parties filed supplemental briefs on May 9, 2011. The parties have also filed additional briefs addressing the

United States Patent and Trademark Office's ("PTO's") ongoing reexamination of all three patents. SRA represents that the PTO has now allowed all asserted claims of the '494 and '352 patents without amendment but has issued a final rejection of the '571 patent.

After consideration of the claims, specification, prosecution history, and other relevant evidence, and after hearing the arguments of the parties, the court now construes the language of the '494 and '352 patents for which the parties have offered different interpretations. Preliminarily, however, the court notes that its practice is normally to hear summary judgment motions that turn on claim construction concurrently with claim construction. The reason is that

> the legal function of giving meaning to claim terms always takes place in the context of a specific accused infringing device or process. While a trial court should certainly not prejudge the ultimate infringement analysis by construing claims with an aim to include or exclude an accused product or process, knowledge of that product or process provides meaningful context for the first step of the infringement analysis, claim construction.

Wilson Sporting Goods Co. v. Hillerich & Bradsby Co., 442 F.3d 1322, 1326-7 (Fed. Cir. 2006). Unfortunately, this practice was not followed here and some of the claim terms were construed outside the context of the infringement and invalidity issues. The constructions may conceivably need revision when the court has a better understanding of the accused methods.

I. U.S. PATENT NO. 5,544,352

The '352 patent, like the other asserted patents, relates to the use of non-semantic link analysis to enhance computerized searching of electronic databases. According to the "Background of the Invention," the claims overcome problems associated with "Boolean" searching of document databases:

> Current computer search programs use a text-by-text analysis procedure (Boolean Search) to scan a database and retrieve items from a database. The attorney must input a string of text, and the computer evaluates this string of text. Then the computer retrieves items from the database that match the string of text . . . . However, Boolean searches of textual material are not very efficient.

'352 patent col.1 ll.27-40.

Addressing these issues,

> The invention begins with an existing database and indexes the data by creating a numerical representation of the data. This indexing technique called proximity indexing generates a quick-reference of the relations, patterns, and similarity found among the data in the database. Using this proximity index, an efficient search for

> pools of data having a particular relation, pattern or characteristic can be effectuated. This relationship can then be graphically displayed.

Id. at col.3 ll.51-58.

> Proximity Indexing indexes the data by using statistical techniques and empirically developed algorithms. The resulting search by an advanced data searching program of the Proximity Indexed data is significantly more efficient and accurate than a simple Boolean search.

Id. at col.4 ll.9-13.

Because the asserted patents share many common terms, the disputed claim terms are construed consistently across the patents. *See NTP, Inc. v. Research in Motion, Ltd.*, 418 F.3d 1282, 1293 (Fed. Cir. 2005).

A. "A non-semantical method"

The parties agree that a "non-semantical method" requires the examination of non-semantical relationships between records (or "objects") in a database. The parties also agree that a "semantical method" accounts for an object's words or phrases. However, defendants seek a negative limitation for "non-semantical" that excludes consideration of words or phrases, thus disclaiming word-based searching. The parties' proposed constructions are as follows:

CLAIM LANGUAGE	SRA'S PROPOSED CONSTRUCTION	DEFENDANTS' PROPOSED CONSTRUCTION
"a non-semantical method"	A method reciting steps that analyze or use non-semantical relationships (*i.e.*, citation relationships)	A method of analysis that does not account for phrases and words in a textual object and that is based on explicit references to other textual objects or A method of analysis that is based on direct relationships between textual objects and that otherwise does not account for phrases and words in a textual object

SRA contends that the claims, specification, and prosecution history of the '352 patent compel the conclusion that semantic methods based on words and phrases are not excluded from non-semantical methods. Claim 26 recites the following:

> 26. A non-semantical method for numerically representing objects in a computer database and for computerized searching of the numerically represented objects in the database, wherein direct and indirect relationships exist between objects in the database, comprising:
>
> marking objects in the database so that each marked object may be individually identified by a computerized search;
>
> creating a first numerical representation for each identified object in the database based upon the object's direct relationship with other objects in the database;
>
> storing the first numerical representations for use in computerized searching;
>
> analyzing the first numerical representations for indirect relationships existing between or among objects in the database;
>
> generating a second numerical representation of each object based on the analysis of the first numerical representation;
>
> storing the second numerical representation for use in computerized searching; and
>
> searching the objects in the database using a computer and the stored second numerical representations, wherein the search identifies one or more of the objects in the database.

'352 patent claim 26. SRA further points to dependent claim 27 which specifically recites the use of "semantic indexing techniques" which account for words and phrases in combination with the non-semantic techniques of independent claim 26:

> 27. The non-semantical method of claim 26, wherein the objects in the database include words, and **semantic indexing techniques are used in combination with the non-semantical method**, **the method further comprising** the step of creating and storing a boolean word index for the words of the objects in the database.

Id. at claim 27 (emphasis added). According to SRA, claim 27 recites that the same "non-semantic method" of claim 26 "further comprises" the creation of a "boolean word index," which necessarily accounts for words or phrases in the database. In response, defendants claim that SRA's position ignores the "non-" in the claim term itself. Defendants contend that the use of the term "comprising" in claim 26 and the addition of semantical steps in dependent claims is not enough to rewrite the plain meaning of the "non-" modifier. *See Dippin' Dots, Inc. v. Mosey*, 476 F.3d 1337, 1343 (Fed. Cir. 2007) (holding that "comprising" does not "render every word and phrase therein open-ended–especially where, as here, the patentee has narrowly defined the claim term it now seeks to have broadened").

In its supplemental briefing, SRA also contends that this court should give significant weight to SRA's arguments submitted to the PTO in the co-pending reexamination of all the asserted claims. The Federal Circuit has instructed courts to "monitor" ongoing reexamination or prosecution activity "to ascertain whether [the court's] construction of any of the claims has been impacted by further action at the PTO or any subsequent proceedings." *Proctor & Gamble Co. v. Kraft Foods Global, Inc.*, 549 F.3d 842, 848-49 (Fed. Cir. 2008). SRA also points out that the panel handling the reexamination of the '352 patent has allowed all of the asserted claims of the '352 patent without amendment over the prior art that defendants submitted to the PTO. Dkt. No. 373 Ex. 1 at pg. 3. SRA cites to *Dura Global Techs., Inc. v. Magna Donnelly Corp.*, 2010 WL 4259615, at *16-18 (E.D. Mich. Oct. 25, 2010) for the proposition that the court should rely on statements made in co-pending reexamination proceedings, even if they are for the benefit of the plaintiff. In *Dura Global*, the court relied on statements made by the plaintiff in a reexamination proceeding that a term did not "exclude pulleys" to reject defendant's construction that required a negative limitation excluding "pulleys" in the co-pending litigation. *Id.* The court went on to note that the plaintiff's statements in reexamination did not appear to be self-serving because it was the court's technical advisor, not the plaintiff, who brought the reexamination statements to the court's attention. *Id* at 17.

In response, defendants argue that because SRA's reexamination statements were filed with the PTO during litigation, they are merely self-serving and should be accorded little, if any, weight. *See, e.g., Moleculon Research Corp. v. CBS, Inc.*, 793 F.2d 1261, 1270 (Fed. Cir. 1986) (noting that papers filed with the PTO during litigation "might very well contain merely self-serving statements which likely would be accorded no more weight than testimony of an interested witness or argument of counsel").

As the Federal Circuit explained in *Springs Window Fashions LP v. Novo Indus., L.P.*, "[t]he public notice function of a patent and its prosecution history requires that a patentee be *held to what he declares* during the prosecution of his patent." 323 F.3d 989, 995 (Fed. Cir. 2003) (emphasis added). Applying this principal, the Federal Circuit has been willing to bind a patentee to its statements made during prosecution. *See, e.g., id.* (holding patentee to prosecution statements requiring that cutters be "independently movable"). It does not follow from the use of a patentee's

statements against the patentee, however, that a patentee may necessarily use statements made during co-pending reexamination for its own benefit. *See Mondis Tech., Ltd. v. Hon Hai Precision Indus. Co. Ltd.*, No. 07-565, 2011 U.S. Dist. LEXIS 7140, at *94 (E.D. Tex. Jan. 24, 2011) ("The statements to which Plaintiff refers were made in re-examination merely months before Plaintiff filed its claim construction brief–thus they should be given little, if any, weight."). Here, SRA's proposed construction is included verbatim in SRA's statements to the PTO. Unlike the plaintiff in *Dura Global*, SRA informed the court of its reexamination arguments within weeks of making them to the PTO. Dkt. No. 351. SRA's reliance on *Adams Respiratory Therapeutics, Inc. v. Perrigo Co.*, 616 F.3d 1283, 1287 (Fed. Cir. 2010) is misplaced. In that case, the Federal Circuit simply found that the defendant was incorrect in arguing that the plaintiff had expressly defined "equivalent as meeting all of the requirements of the FDA's bioequivalence guidelines" during reexamination. *Id.* The *Adams Respiratory* court did not adopt a verbatim construction advanced by the plaintiff during reexamination. To be sure, the panel of examiners handling the reexamination of the '352 patent issued a Final Office Action, wherein they allowed all of the asserted claims in the '352 patent. Dkt. No. 373 Exh. 1. But prosecution is not a platform for a patentee to "enlarge, diminish, or vary the limitations in the claims." *Markman v. Westview Instruments, Inc.*, 52 F.3d 967, 980 (Fed. Cir. 1995). SRA fails to identify anything in the reexamination of the '352 patent indicating that the examiners adopted SRA's proposed construction of "non-semantical." Therefore, SRA's self-serving statements made in the concurrent reexamination proceedings are entitled to little weight.

The court turns to the claims and the specification, which the Federal Circuit has identified as "the single best guide to the meaning of a disputed term," *Phillips v. AWH Corp.*, 415 F.3d 1303, 1321 (Fed. Cir. 2005). At first glance, the term "non-semantical method" seems to exclude consideration of any semantical relationships. SRA argues that a complete exclusion of semantical considerations is irreconcilable with the fact that the steps of claim 26 require numerically representing textual citations contained in documents, and, therefore, necessarily account for the words and phrases of the citations in those documents. However, defendants' latest construction takes into account phrases and words in textual objects only as necessary to analyze direct

relationships. Therefore, SRA's argument that defendants' proposed construction does not account for words and phrases at all is not correct.

Dependent claim 27 recites the "non-semantical method . . . the method further comprising the step of creating and storing a boolean word index." '352 patent claim 26. According to SRA, this language inherently recognizes that the non-semantical method of claim 26 can include semantical techniques – namely, a boolean word index. SRA's argument is unconvincing. Regardless of the addition of semantic techniques in claim 27, the steps of claim 26 are expressly qualified as "non-semantical." In *Regents of Univ. of Cal. v. DakoCytomation Cal., Inc.*, the Federal Circuit noted that "while it is true that dependent claims can aid in interpreting the scope of claims from which they depend, they are only an aid into interpretation and are not conclusive." 517 F.3d 1364, 1375 (Fed. Cir. 2008). The Federal Circuit went on to hold that the term "unique sequences" excluded "repetitive sequences," even though certain dependent claims did not exclude repetitive sequences. *Id.* at 1376. Similarly, independent claim 26 excludes semantic techniques, even though dependent claim 27 adds a semantic technique. In fact, claim 27 states that "semantic indexing techniques are used *in combination with* the non-semantical method," thereby demonstrating that the two are indeed separate methods. '352 patent col.35 ll.54-60 (emphasis added).

Moreover, the term "comprising" cannot be used to render every word or phrase open-ended. Dependent claim 27 specifically recites the "non-semantical" method of claim 26 as "further comprising" a semantic technique, but that does not necessarily mean that the non-semantical elements of claim 26 can be re-written to include semantic techniques. SRA's construction impermissibly stretches the bounds of the term "comprising."

SRA's reliance on the doctrine of claim differentiation is similarly misplaced, especially where the patentee distinguished prior art as allegedly disclosing semantic rather than non-semantical techniques. Dkt. No. 284-5 at 3 ("In contrast, the novel search techniques claimed in the above-titled application do not rely on the semantic model approach and are not described in any of the references."). To be sure, SRA contends that other statements made during the prosecution of the '352 patent specifically combine semantic and non-semantic techniques:

> Although the novel techniques [*i.e.*, non-semantical methods] may be used in combination with semantical techniques, they may also be conducted without the use

> of any semantical techniques. Use of the novel methods with and without semantic techniques is claimed.

See Dkt. No. 314-1 at 29. But this statement actually supports defendants' construction, as it is consistent with claim 26 as encompassing the novel non-semantic techniques and claim 27 as encompassing the combination of novel non-semantic techniques with semantic techniques. The statement does not explicitly show that semantic techniques are part of, rather than additions to, the claimed non-semantical method. As a whole, the prosecution history supports dependants' proposed construction.

The specification also supports defendants' proposed construction. SRA is correct in noting that the specification describes a Proximity Indexing Application Program (*i.e.*, the program that "numerically represents" objects) that conducts semantic proximity indexing techniques in addition to non-semantical indexing. *See* '352 patent col.8 ll.33-42. However, these semantical techniques are separate from the non-semantical method of claim 26. *See id.* at col.8 ll.33-36 (noting that the "sequence of subroutines that comprise the Boolean Indexing Subroutine [] are executed during another stage of the Proximity Indexing Application Program"). SRA also points to parts of the specification that describe pool searches using (1) non-semantical matrices (the Opinion Citation Matrix and the Opinion Similarity Matrix) and (2) "other matrices" to rank the importance of objects in the pool as well as determine the similarity of the objects in the pool. *Id.* at col.20 l.65-col.21 l.3 (pool-similarity searching); *id.* at col.21 ll.26-30 (pool-importance searching). At the same time, the specification discloses pool searches of the Computer Search Program for Data Represented in Matrices ("CSPDM") that "evaluate the Section Word Similarity Matrix as well as other matrices to determine whether or not to retrieve a full textual object." *Id.* at col.19 ll.13-15. The "Section Word Similarity Matrix" is a semantic word-based matrix. *See id.* at col.19 ll.6-12. According to SRA, these parts of the specification disclose using both semantic and non-semantical based matrices in performing pool searches. Notably, these steps are directed to dependent claims 39 and 40. Each of those claims add pool-similarity and pool-importance limitations to the searching step of the non-semantical method of claim 26. *Id.* at claims 39 and 40. But the addition of semantic steps in later claims does not convert the "non-semantical method for numerically representing objects . . . and computerized searching" into a semantic one. *Cf. TiVo, Inc. v. Echostar Comm'ns Corp.*, 516 F.3d

1290, 1302 (Fed. Cir. 2008) (dependent claim reciting steps occurring after steps of independent claim did not inform construction of independent claim).

In view of the claims and the specification, the term "non-semantical method" is construed as "a method of analysis that is based on direct relationships between textual objects and that otherwise does not account for phrases and words in a textual object." The court thus accepts defendants' revised proposed construction of "non-semantical."

B. "objects" and "of each object"

The parties also dispute the meaning of the terms "object" in multiple limitations of claims 26, 35 and 41 of the '352 patent. Defendants contend that the claim term "object" refers to every object that may exist in the database, while SRA argues that "object" refers to the set of objects to be numerically represented as referenced in the preamble. The parties' proposed constructions are as follows:

CLAIM LANGUAGE	SRA'S PROPOSED CONSTRUCTION	DEFENDANTS' PROPOSED CONSTRUCTION
"A . . . method for numerically representing objects in a computer database and for computerized searching of numerically represented objects in the computer database"	A computer-implemented . . . method for numerically representing a set of objects in a computer database and for computerized searching of the set of numerically represented objects in the computer database.	No construction necessary.
"generating a second numerical representation of each object based on the analysis of the first numerical representation"	generating a second numerical representation of each identified object within the set of numerically represented objects based on the analysis of the first numerical representations	generating a second numerical representation of each object in the database that describes the indirect citation relationships found by analyzing the first numerical representation

As an initial matter, SRA's proposed construction adds the language "computer-implemented" to the preamble. As defendants point out, the plain claim language already informs that a computer is required by certain aspects of the claimed method, as it expressly refers to "a computer database," "computerized searching," and "using a computer." But SRA contends that the specification and prosecution history demand that all of the claimed steps be computer implemented.

SRA points to multiple instances during the prosecution of the patent where prior art references were distinguished as pertaining to "manual" editing and clustering while the '352 patent claims all pertained to computerized indexing. *See* Dkt. No. 246-6 at 32. Moreover, the specification notes that any manual methods are "prone to error" while the claimed invention "can be recognized without direct human intervention." '352 patent col.2 ll.25-42 and col.7 ll.17-28. However, adding "computer-implemented," which is found nowhere in the specification, essentially renders the other computer-specific language superfluous. *See Mangosoft, Inc. v. Oracle Corp.*, 525 F.3d 1327, 1330-31 (Fed. Cir. 2008) (rejecting a construction that would render claim terms superfluous). Accordingly, the court finds that SRA's proposed addition of "computer-implemented" is unwarranted.

Next, the court turns to SRA's proposal to add "a set of" to the claim. Defendants accuse SRA of rewriting the claim by adding the language. According to defendants, the preamble requires that *all* objects in the database must be numerically represented. SRA relies on *Lava Trading, Inc. v. Sonic Trading Management, LLC*, 445 F.3d 1348 (Fed. Cir. 2006) for the proposition that where a preamble uses a singular article, the claim language supports coverage of a subset rather than an entire set. *See id.* at 1354 ("By selecting the word 'a' instead of 'all,' the Applicant set forth a method wherein the traders may request and receive information for only a subset of the securities (i.e., one or more) . . . ."). In this case, the preamble of claim 26 uses the plural terms "objects" and "numerically represented objects," which defendants argue must necessarily encompass all objects. In addition, defendants contend that the preferred embodiment uses "each full textual object (*e.g.*, whole judicial opinion, statute, etc.)" as the starting point for creating numerical representations. '353 patent col.4 ll.27-31. According to defendants, the specification nowhere describes assigning numerical representations to only a portion of full textual objects.

The court finds that both the literal language of the claim and the disclosed embodiments compel an interpretation that "numerically representing objects in a computer database" does not require every object in the database to be numerically represented. Critically, the claim language only refers to "objects," not "all objects." The fact that the term is plural is not enough to require the inclusion of every single object. *See Versa Corp. v. Ag-Bag Int'l Ltd.*, 392 F.3d 1325, 1330 (Fed.

Cir. 2004) (holding that the plural form describes "a universe ranging from one to some higher number" rather than necessarily all). Moreover, the specification describes several embodiments that numerically represent less than all of the objects in the database. While the preferred embodiment does numerically represent all full textual objects, it does not numerically represent the relationships of page objects, phrase objects, non-textual objects, or other subset objects that exist in the database, or even the marked word objects. '353 patent col.11 l.61-col.12 l.24. Claim 26 recites "objects" rather than "full textual objects," and therefore refers to objects generally rather than full textual objects only. As SRA points out, the '352 patent defines objects as including portions of objects as well: "Textual objects other than full textual objects may be subsets of full textual objects and of each other. For example, a section, page, or paragraph of text taken from a longer text may be treated as a textual object." *Id.* at col.12 ll.18-23. Furthermore, the patent describes providing non-semantic numerical representations for only a pool of objects (which is less than all objects in a database) in the paradigm routines. *Id.* at col.26 ll.40-46 (noting a paradigm proximity matrix including only a pool of "k" objects).

Along these same lines, the preamble provides the antecedent basis for the term "objects" in the body of the claims. Because the "objects" in the body of the claim are the "objects" that will be "numerically represented" in the preamble, and because the term "objects" in the preamble is not necessarily all objects in the database, the "objects" in the body of the claim does not necessarily mean all objects in the database either. *See Lava Trading*, 445 F.3d at 1353-55. Accordingly, the court accepts SRA's proposed construction of "objects" and "of each object," but declines to insert the language "computer-implemented."

C. "identified object"

The parties also disagree as to whether an "identified object" as recited in claim 26 of the '352 patent is a "marked object" (as defendants contend) or "an object identified by a search using a computer and the second numerical representation" (as plaintiff contend). The parties' proposed constructions are as follows:

CLAIM LANGUAGE	SRA'S PROPOSED CONSTRUCTION	DEFENDANTS' PROPOSED CONSTRUCTION
"identified object"	An object identified by a search using a computer and the second numerical representation	each object that has been marked

SRA points to the language of claim 26 reciting "marking objects in the database so that each marked object *may be* individually identified by a computerized search." '352 patent col.35 ll.33-35 (emphasis added). According to SRA, this language precludes defendants' construction equating "marked objects" with "identified objects." Furthermore, SRA notes that the final portion of claim 26 explains how the "search identifies one or more of the objects in the database" and the search is performed "using a computer and the stored second numerical representation." *Id.* at col.35 ll.50-53. SRA contends that this claim language shows that "identified objects" are those "identified by a search using a computer and the second numerical representation."

In support of their proposed construction, defendants point to step two of claim 26, which refers to "creating a first numerical representation for each identified object." *Id.* at col. 35 ll.36-39. According to defendants, the "identified object" in step two can only be referring to the object marked in the first step, otherwise, "each identified object" would lack antecedent basis. Put differently, defendants claim that SRA's construction is illogical because claim 26 explicitly requires that the "identified object" exist before a computerized search is conducted, but SRA's construction defines the "identified object" as the result of a computer search.

35 U.S.C. § 112 ¶ 2 requires an applicant to particularly point out and distinctly claim that which is regarded as the invention. But the issue of antecedent basis typically arises when dealing with a definite article ("the" or "said"). And while the convention is useful, it is not mandatory. *See In re Skvorecz*, 580 F.3d 1262, 1268-69 (Fed. Cir. 2009) (citing MPEP § 2173.05(e) for the proposition that "the failure to provide explicit antecedent basis for terms does not always render a claim indefinite"). Here, the term at issue is not "the identified object," but rather, "each identified object." Therefore, the lack of antecedent basis, while perhaps confusing, does not force defendants' proposed construction.

Even more to the point, defendants fail to address the fact that under their construction, the preferred embodiment would be excluded. The Federal Circuit has cautioned that "[a] claim construction that excludes a preferred embodiment from the scope of the claim is rarely, if ever, correct." *MBO Labs., Inc. v. Becton, Dickinson & Co.*, 474 F.3d 1323, 1333 (Fed. Cir. 2007). Here, the disclosed embodiments contemplate that not every object in the database (marked, identified, or otherwise) will receive a second numerical representation. Defendants fail to address that none of the disclosed embodiments creates numerical representations for marked word objects and the numerical representations for marked paragraphs is "optional." '352 patent col.16 l.41-col.17 l.25. Defendants' proposed construction is therefore problematic because the final step of claim 26 states that a "search identifies one or more of the objects in the database" and the search is performed "using a computer and the stored *second numerical representation*." *Id.* at claim 26 (emphasis added). Put differently, not all marked objects receive a second numerical representation, yet only objects that have received a second numerical representation are identified by the search. Contrary to the disclosed embodiments, defendants' proposed construction would require a numerical representation for every marked object. Accordingly, the court accepts SRA's proposed construction of "identified object."

D. "analyzing the first numerical representation for indirect relationships"

While SRA contends that no construction is necessary for "analyzing the first numerical representation for indirect relationships" in claim 26, defendants argue that construction for this term is necessary to avoid potential confusion over the scope of the claim. In its supplemental briefing, SRA submitted an alternative construction. The proposed constructions are as follows:

CLAIM LANGUAGE	SRA'S PROPOSED CONSTRUCTION	DEFENDANTS' PROPOSED CONSTRUCTION
"analyzing the first numerical representation for indirect relationships"	No construction necessary or using the first numerical representation to mathematically account for indirect relationships	using the first numerical representation to locate and identify the indirect relationships

Defendants point to the operations of pattern creation for their support of "locating and identifying" in the specification. The specification describes using numerical representations of direct relationships (Opinion Citation Matrix) to locate and identify indirect relationships, which are identified by corresponding opinion pattern vectors. '352 patent col.15 ll.23-28. According to defendants, this process occurs in each disclosed embodiment.

The term "analyzing" appears to require "locating" and "identifying."[1] Indeed, claim 26 requires "analyzing the first numerical representation *for* indirect relationships," or put differently, determining whether or not the indirect relationship is present. *Id.* at claim 26 (emphasis added); *see also* col.15 ll.23-28, Fig. 6. SRA contends that during reexamination of the '571 patent, the examiner made statements endorsing SRA's proposed construction of "analyzing for indirect relationships." But SRA confuses the examiners description of "cluster analyzing" from the '571 reexamination proceeding with "proximity analyzing" from the '352 patent. This is particularly inappropriate because the concepts of "cluster links" and "cluster analyzing" were new matter added to the '494 continuation-in-part application (from which the application for the '571 patent depends). But if the court were to consider the reexamination history of the '571 patent, it must consider the history as a whole. Notably, SRA attempted to distinguish the "proximity analysis" limitation of claim 5 of the '571 patent by arguing that "[t]he [prior art] method never actually locates any *particular* path or analyzes or determines in any way the specific indirect relationship between two nodes . . . ." Dkt. No. 437 Exh. 4 at 27. If anything, the reexamination history actually supports defendants' proposed construction.

SRA is correct in noting that dependent claim 32 specifies that claim 26's "step of analyzing" direct relationships for indirect relationships includes "weighing" where "some indirect relationships are weighed more heavily than other indirect relationships." *Id.* at claim 32. This step contemplates quantifying the strength of indirect relationships using a weighted mathematical analysis after the indirect relationships are located and identified by corresponding opinion pattern vectors. *Id.* at

---

[1] During the claim construction hearing, SRA noted that "you have to locate and identify the indirect relationship to quantify it so those are steps that are part of the analysis, they're just not all of the steps encompassed by the analysis." Dkt. No. 372 at 208. SRA has since argued that this was a misstatement and that the '352 patent discloses "analyses that quantify or score indirect relationships without identifying patterns of indirect relationships." Dkt. No. 378 at 3.

col.13 ll.34-55. In other words, "analyzing" can also include weighing and quantifying indirect relationships. Accordingly, the court construes "analyzing the first numerical representation for indirect relationships" to mean using the first numerical representation to at least locate and identify the indirect relationships.

II. U.S. PATENT NO. 5,832,494

The '494 patent is a continuation-in-part of the '352 patent and also describes a "computerized research tool for indexing, searching and displaying data." '494 patent Abstract. As with the '352 patent, the '494 patent discloses the use of non-semantic methods of analyzing relationships between database objects to index them for subsequent searching. *See id.*

The '494 patent also describes "a cluster link generation algorithm which uses links and nodes to index and search a database or network." Representative (and asserted) claim 1 from the '494 patent recites:

> 1. A method of analyzing a database with indirect relationships, using links and nodes, comprising the steps of:
>
> selecting a node for analysis;
>
> generating candidate cluster links for the selected node, wherein the step of generating comprises an analysis of one or more indirect relationships in the database;
>
> deriving actual cluster links from the candidate cluster links;
>
> identifying one or more nodes for display; and
>
> displaying the identity of one or more nodes using the actual cluster links.

A. "cluster link"

The parties agree that a "cluster link" is a relationship between two nodes. However, the proposed constructions differ in two respects: SRA's proposal requires a (1) "mathematical analysis" of (2) "weights of direct links in a set of paths between two nodes" while defendants' proposal requires a (1) "statistical analysis" of (2)"multiple relationships between nodes in a database." The parties' proposed constructions are as follows:

CLAIM LANGUAGE	SRA'S PROPOSED CONSTRUCTION	DEFENDANTS' PROPOSED CONSTRUCTION
"cluster link"	the relationship defined by mathematically analyzing weights of direct links in a set of paths between two nodes	a relationship between two nodes based upon a statistical analysis of multiple relationships between nodes in a database

Where the "patentee has chosen to be his or her own lexicographer by clearly setting forth an explicit definition for a claim term," that construction governs. *Johnson Worldwide Assocs., Inc. v. Zebco Corp.*, 175 F.3d 985, 990 (Fed. Cir. 1999). Defendants contend that the patentees became their own lexicographers by defining the term "cluster link" in the glossary of the "V-Search Publisher's Toolkit User's Manual" ("V-Search Manual"), which the patentees incorporate by reference into the '494 patent. *See* '494 patent col.1 ll.18-22; V-Search Manual at 152. The V-Search Manual contains the following definition:

> **Cluster link**: A relationship between two nodes based, by default, on a statistical analysis of multiple relationships between two nodes in a database. For example, two nodes, both directly linked to the same intermediate nodes, may be indirectly linked through many paths and therefore have a Cluster link between them.

The V-Search Manual is relevant to claim construction. *See, e.g., Cook Biotech Inc. v. Acell, Inc.*, 460 F.3d 1365, 1375-77 (Fed. Cir. 2006) (construing claims based on a patent incorporated by reference). However, SRA contends that this incorporated definition refers only to a particular commercial embodiment of the invention and is inconsistent with the specification.

As the Federal Circuit acknowledged in *Phillips*, "there is sometimes a fine line between reading a claim in light of the specification, and reading a limitation into the claim from the specification." *Phillips v. AWH Corp.*, 415 F.3d 1303, 1323 (Fed. Cir. 2005) (en banc) (quoting *Comark Commc'ns, Inc. v. Harris Corp.*, 156 F.3d 1182, 1186-87 (Fed. Cir. 1998)). In this case, the '494 patent does address "weights" and "direct links" as highlighted in SRA's proposed construction. For example, the specification refers to a cluster link generator that "generates a set of cluster links based upon both the direct links and the indirect paths." '494 patent col.21 ll.57-63. Also, the specification describes the strength of implied relationships as "depend[ing] on the length of the path, and on the weights of the individual direct links along the path." *Id.* at col.22 ll.11-15.

All of these statements, however, refer to specific examples of cluster links or cluster link generators. More broadly, the specification states that:

> A cluster link is a relationship between two nodes, for example, two nodes both directly linked to the same intermediate nodes, may be indirectly linked through many paths and therefore have a cluster link between them. The cluster links may be determined using the specific or general methods described later for finding relationships in a database. However, the preferred method is through using a proximity Indexing Application Program.

Id. at col.13 ll.24-33. Cluster links may even be established manually. *Id.* at col.13 ll.21-24. And while the specification does show that a cluster link can be defined by two nodes directly linked, *see* '494 patent col.8 ll.52-57, a cluster link is not necessarily limited to the analysis of direct relationships. In fact, "a relationship may be semantical, non-semantical, stated, implied, direct, indirect, statistical, and/or theoretical." *Id.* at col.13 ll.1-3. The V-Search Manual's definition highlighting the "analysis of multiple relationships between two nodes," which is itself part of the specification, is supported by the specification as a whole.

The specification further explains that cluster links are relationships based upon statistical analysis. For example, the specification explains that the "object of the invention" is to "utilize statistical techniques" to enhance research. '494 patent col.7 ll.35-43, col.11 ll.35-37, col.50 ll.22-26. SRA does not point to any portion of the specification referring to anything other than a "statistical analysis." And while SRA argues that the nodes need not be "in a database," the plain language of the asserted claims all refer to a "database."

The PTO has recently allowed all the asserted claims in the '494 patent. Dkt. No. 422-2. In the Reasons for Allowance in the Final Office Action for the '494 patent, the examiner explained why the asserted claims are distinguishable from the prior art:

> With regard to the rejections utilizing Lucarella, patent owner (PO) has shown that Lucarella discloses the use of concept links between objects based on analyzing a group of objects based on a query. Therefore Lucarella does not teach generating a candidate cluster link set, the cluster links being functions of the weights associated with the direct links and the weight of the previous cluster link, (see Fig. 3G of the Egger Patent and the corresponding discussion) from which actual cluster links are derived.

Id. at 3. SRA contends that the examiner's statement effectively recognizes that "cluster links" must be "functions of the weights associated with direct links," thereby explicitly adopting SRA's view that the analysis must pertain to the weights of direct links. The court is not convinced that the examiner's statement is a clear endorsement of plaintiff's proposed construction of "cluster link." Indeed, nothing in the examiner's statement is inconsistent with defendants' proposed construction.

The term "functions" mentioned by the examiner is not inconsistent with a "statistical analysis." Significantly, the specification discloses only statistical techniques, and SRA cites to nothing in the specification or in the reexamination prosecution history where the term "mathematical" appears. In addition, the "analysis of multiple relationships" in defendants' proposed construction encompasses an analysis of the "previous cluster link" mentioned by the examiner. At bottom, the reexamination history indicates that the examiner was merely stating his reasons for allowance and was not providing a definition for "cluster link."

SRA also contends that the court is bound to resolve ambiguities with respect to claim construction in a manner that preserves the validity of the patent. According to SRA, defendants seek a broad interpretation of "cluster link" in order to show that the invention is disclosed in the prior art. SRA suggests that the court is bound to accept a claim construction that avoids the prior art. The issue of invalidity is not before the court at this time. But in any event, claims are to be construed using intrinsic and pertinent extrinsic evidence that bears on the meaning of terms as used in the patent claims. The basic framework of claim construction does not accommodate modification of claim language simply to avoid prior art.[2] The maxim that claims should be construed to avoid invalidity is limited to cases where the evidence is inconclusive. *See Phillips*, 415 F.3d at 1327 ("[W]e have limited the maxim [that claims should be construed to preserve their validity] to cases in which 'the court concludes, after applying all the available tools of claim construction, that the claim is still ambiguous.'"); *see also Saunders Group, Inc. v. Comfortrac, Inc.*, 492 F.3d 1326, 1335 (Fed. Cir. 2007) ("[W]e hold only that the court's validity analysis cannot be used as basis for adopting a narrow construction of the claims.").

Both the specification and the V-Search Manual (which is incorporated into the specification by reference) describe cluster links as a relationship between two nodes based upon a statistical analysis of multiple relationships between nodes in a database. Accordingly, the court accepts defendants' proposed construction of "cluster link."

B. "candidate cluster link"

The parties' proposed constructions for "candidate cluster link" are as follows:

---

[2] *See* Peter Menell et al., *Patent Claim Construction: A Modern Synthesis and Structured Framework*, 25 Berkeley Tech. L.J. 711, 765 (2010).

CLAIM LANGUAGE	SRA'S PROPOSED CONSTRUCTION	DEFENDANTS' PROPOSED CONSTRUCTION
"candidate cluster link"	a set of cluster links from a selected node to other nodes from which actual cluster links may be derived	the set of all possible cluster links between a search node and a target node

The patentees twice define "candidate cluster links" as "the set of all possible cluster links between a search node and a target node." '494 patent col.21 l.66-col.22 l.1; col.22 ll.53-55.[3] The specification teaches that the term "candidate cluster links" may be broader than the *subset* of candidate cluster links from which actual cluster links are derived. *Id.* at col.21 l.64-col.22 l.4 ("In this embodiment, only a subset of the candidate cluster links, the actual cluster links, which meet a certain criteria are used to locate nodes for display."). While portions of the specification refer to examining only a set of cluster links (*see id.* at col.24 ll.12-14), these portions do not relate to the claimed step of "generating candidate cluster links." Rather, they refer to the step of "deriving actual cluster links from candidate cluster links" as recited in claim 1.

SRA contends that the patent also refers to "start nodes" and therefore, "candidate cluster links" should not be limited to "search nodes" and "target nodes" only. SRA points to a requested amendment in the prosecution history that would have replaced "search node" with "start node," but was never entered. Defendants, however, do not dispute that a "search node" may encompass a "start node." Dkt. No. 435 at 3.

SRA also argues that the PTO adopted its proposed construction for "candidate cluster links" during reexamination. In explaining a reason for allowance, the examiner stated that "Lucarella does not teach generating a candidate cluster link set . . . from which actual cluster links are derived." Dkt. No. 422-2 at 3. According to SRA, this statement adopts SRA's argument that "candidate cluster links" are the "set of cluster links . . . from which actual cluster links are derived." SRA's argument is unconvincing. The examiner was not adopting any definition of "candidate cluster link," but rather, was merely paraphrasing a claim limitation. Defendants do not dispute that

[3] Defendants admit that "all possible cluster links" does not mean "unlimited." Instead, the number of all possible cluster links between a search node and a target node is dictated by the relationship between the nodes. For example, defendants recognize that Figure 3G of the '494 patent shows only one candidate cluster link between N1 and N3 (C3).

actual cluster links are derived from candidate cluster links. Indeed, the claims specifically recite this. *See* '494 patent claim 1 ("deriving actual cluster links from the candidate cluster links"). Nothing in the examiner's reexamination statement is inconsistent with the '494 patent's express definition of a "candidate cluster link" as "the set of all possible cluster links between a search node and a target node."

Because it conforms to the explicit definition in the specification, the court accepts defendants' proposed construction of "candidate cluster link."

C. "wherein the step of generating comprises an analysis of one or more indirect relationships in the database"

The parties' proposed constructions are as follows:

CLAIM LANGUAGE	SRA'S PROPOSED CONSTRUCTION	DEFENDANTS' PROPOSED CONSTRUCTION
"wherein the step of generating comprises an analysis of one or more indirect relationships in the database"	No construction is necessary	wherein the step of generating comprises identifying and classifying one or more non-semantical relationships that are characterized by at least one intermediate node between two nodes in the database

The parties agree that an "indirect relationship" is a relationship where at least one intermediate object exists between two objects and where the intermediate objects connect the two objects through a chain of citations.

The specification describes a cluster link generation algorithm that "retrieves a list of URLs and classifies them as the direct links to be analyzed." '494 patent col.49 ll.58-62. The cluster link generator also traces these direct links to their destination nodes and classifies them "as being indirectly linked to the source node, and the links to these nodes are added to the list of candidate cluster links." *Id.* at col.49 l.65-col.50 l.2. In other words, the cluster link generation algorithm identifies and classifies indirect relationships. This process is reflected in defendants' proposed construction. However, the patent also describes quantifying the strength of paths made of indirect relationships when generating candidate cluster links. *Id.* at col.22 ll.5-15 ("The strength of the implied relationship depends on the length of the path, and on the weights of the individual direct links along the path."). Accordingly, the court construes "analysis of one or more indirect relationships in the database" to mean at least identifying and classifying one or more non-

semantical relationships that are characterized by at least one intermediate node between two nodes in the database.

D. "selecting a node for analysis"

The parties' proposed constructions are as follows:

CLAIM LANGUAGE	SRA'S PROPOSED CONSTRUCTION	DEFENDANTS' PROPOSED CONSTRUCTION
"selecting a node for analysis"	No construction is necessary	identifying, by an end user, a node to be non-semantically analyzed

The patentees state that "[t]he invention pertains to computerized research tools." '494 patent col.1 l.11. The specification also underscores, in multiple instances, that a researcher will use these tools. *See id.* at col.2 ll.46-47, col. 2 ll.57-58, col.2 ll.62-63, and col.3 ll.3-4. Moreover, the specification repeatedly uses the terms "identify," "select," and "choose" interchangeably, explaining that an "end user" or "researcher" performs the identification, selection, or choosing. *See id.* Figures 4A-4I.

SRA argues that "selecting" can be performed by a computer. SRA contends that the specification teaches selecting a node for analysis during indexing and, therefore, before the end user is involved. *Id.* at col.23 ll.43-60. But regardless of whether the required analysis refers to computer actions performed during indexing, the specification makes clear that a researcher selects the object for analysis. SRA's example relates to the candidate cluster link algorithm, which according to Fig. 14B performs its function after a "source node" has been selected. *See Id.* at Fig. 14B, cols.49-50. This "source node" for analysis is the end user's node, as "the end user may monitor sites to which his site (the source web page) is referring, and the end user may view the sites which are referring to his site." *Id* at col.51 ll.5-8.

In multiple places, the '494 patent is clear that "selecting a node" occurs independently from indexing and requires user input. *See id.* at Figs. 4A-4I. The specification explains that:

> The researcher can also identify the type of search to be performed on the selected textual object by selecting the appropriate search in the Analysis box.

Id. at col.35 ll.28-30.

The options in the analysis box enable a researcher to select a textual object outside

the pool of textual objects.

Id. at col.35 ll.42-43.

The researcher must select a single textual object.

Id. at col.25 ll.8-10, 31-34, 35-37.

The researcher must select a pool of full textual objects.

Id. at col. 25 ll.53-62; col.26 ll.5-8, 20-23, 30-36.

At the same time, the claims refer to "displaying" nodes generated after analysis of the selected node, further suggesting that an end user selected the node for analysis. *See id.* at claim 1, 12 and 18.

The court thus accepts defendants' proposed construction of "selecting a node for analysis."

III. U.S. Patent No. 6,233,571

The '571 patent is a divisional of the '494 patent and therefore shares the same specification. However, the PTO has now issued a final rejection of the claims at issue in the '571patent. In light of this rejection, there is no reason to interpret the language in the '571 patent that is the subject of the parties' dispute.

IV. ORDER

For the foregoing reasons, the court construes the claim language for which the parties have offered different interpretations as follows:

CLAIM LANGUAGE	CONSTRUCTION
"a non-semantical method"	A method of analysis that is based on direct relationships between textual objects and that otherwise does not account for phrases and words in a textual object.
"A . . . method for numerically representing objects in a computer database and for computerized searching of numerically represented objects in the computer database"	A . . . method for numerically representing a set of objects in a computer database and for computerized searching of the set of numerically represented objects in the computer database.
"generating a second numerical representation of each object based on the analysis of the first numerical representation"	generating a second numerical representation of each identified object within the set of numerically represented objects based on the analysis of the first numerical representations

CLAIM LANGUAGE	CONSTRUCTION
"identified object"	an object identified by a search using a computer and the second numerical representation
"analyzing the first numerical representation for indirect relationships"	using the first numerical representation to at least locate and identify the indirect relationships
"cluster link"	a relationship between two nodes based upon a statistical analysis of multiple relationships between nodes in a database
"candidate cluster link"	the set of all possible cluster links between a search node and a target node
"wherein the step of generating comprises an analysis of one or more indirect relationships in the database"	wherein the step of generating comprises at least identifying and classifying one or more non-semantical relationships that are characterized by at least one intermediate node between two nodes in the database
"selecting a node for analysis"	identifying, by an end user, a node to be non-semantically analyzed

DATED: 8/31/2011

Ronald M. Whyte

RONALD M. WHYTE
United States District Judge